

From: "Collins, Francis (NHGRI)" <francisc@exchange.nih.gov>
To: "Varmus, Harold" <hv2b@nih.gov>
Subject: FW: new TIGR Data release policy
Date: Fri, 18 Dec 1998 15:53:50 -0500
MIME-Version: 1.0

Uh-oh. See attached message from David Lipman. NIH and DOE will have to decide whether to protest. The Bermuda agreement from 2/98 explicitly indicated that genome sequence on ALL organisms should be immediately released into the public domain. TIGR's plan is not consistent with that. Presumably they will justify this action on the basis of the concern about wholesale downloading and resale of their databases.

This messy outcome could have been prevented by insisting on the same written conditions of grant award that we require of human genome sequencers, but I don't think this has been done by DOE or NIAID for microbial genomes.

I've also forwarded this to Maria, and I'll talk to Ari and Tony to get their views.

Francis

-----Original Message-----

From: David Lipman [<mailto:lipman@ncbi.nlm.nih.gov>]
Sent: Friday, December 18, 1998 11:28 AM
To: Ari.Patrinios@oer.doe.gov; Collins, Francis (NHGRI);
AFAUCI@flash.niaid.nih.gov
Subject: new TIGR Data release policy

Dear Tony, Francis, and Ari,

As you know, DOE and NIH are funding TIGR to sequence a number of microbial genomes and up until recently TIGR made the unfinished sequence data available for ftp on their web site. NCBI had worked with TIGR and other sequencing centers to make unfinished genome data BLAST-searchable on the NCBI site. This site has been quite useful to the community as the users need not search an increasing number of different sites to find interesting homologs. TIGR has just changed this policy and now requires a licensing agreement in order to download the data [text attached]. This licensing agreement makes it essentially impossible for us to download the data and provide the "one-site-blast" service. The restrictions imposed by the agreement will eliminate a number of useful applications of this pre-publication data that other scientists have been providing. Furthermore I believe this policy may influence other genome centers to adopt the same approach. The biomedical community will not be well-served if each scientist must hunt down a number of different sites to find their genome. Note that a microbial genome might be in the "unfinished" or "pre-publication" state for quite a while and that this information is

incredibly useful to the community. Furthermore I think there are a number of other intellectual property issues raised by this new policy (e.g. TIGR selling access to that data to commercial customers).

Although TIGR is still providing unrestricted downloading of the plasmodium sequence data, the downloadable data is much older than what is available for blast searching. So this too is a problem.

I spoke with Claire Fraser, who is now the head at TIGR, to see if something could be worked out and I've also followed up with an e-mail message. I'm not sanguine that anything will happen soon if at all because they have ignored most of our e-mail messages on this subject. I'm not sure what you can do about the problem but I thought it important for you all to be aware of the situation. Please contact me if you have any questions or ideas.

--David

David J. Lipman, MD
Director, NCBI/NLM
National Institutes of Health

Free License for Unfinished Genomic Sequence Data

The Institute for Genomic Research (TIGR) releases data very rapidly to ensure that our scientific colleagues have access to information that may assist them in the search for genes and their biological function. Data releases do not constitute scientific publication, but rather provide investigators with information that may jump-start biological experimentation.

Our incomplete genomes are available for free to the nonprofit research community in advance of publication. The files containing our pre-publication data are large multi-FASTA files containing all sequences grouped into contigs by TIGR Assembler. Upon sending in this license agreement, TIGR will send you instructions on how to transfer the file or files requested. Note that this data may contain errors or be incomplete and should be regarded as preliminary.

The purpose of this agreement is to protect the interests of the TIGR

scientists who are sequencing and annotating these unfinished genomes. Our goal is to finish and publish genomes as rapidly as possible. TIGR intends to publish for each of these genomes a complete analysis such as appears in Fleischmann et al. (Science 269 (1995): 496-512), Fraser et al. (Science 270 (1995): 397-403), and subsequent genome papers. Our policy is that early release should aid the progress of science, but it is not intended to allow others to preempt TIGR's scientific work by rushing to publication in advance of our own efforts.

In return for free access to this data, we ask that you agree to the following conditions:

1. You agree not to re-publish the data in any form, including displaying it on the Web or other electronic forums.
2. You agree not to redistribute the data. TIGR will remain the sole source of this data until publication, at which time TIGR will make the finished data freely available from its own Web site and from other public electronic sources (as it has in the past).
3. If you submit an article for publication that uses any part of the information in the unpublished genomic sequence data, you agree to cite TIGR in both the Materials and Methods and Acknowledgement sections as follows: "Preliminary sequence data was obtained from The Institute for Genomic Research through the website at <http://www.tigr.org>." In addition, the Acknowledgements should contain the following text: "Sequencing of [project name] was accomplished with support from [name of funding agency]." The name of the funding agency for each project can be found on the TIGR website. If your submission is accepted, you also agree to notify us by email at the time of acceptance, and to send us a pre-print of the article prior to publication.

You may email this agreement to license@tigr.org, or fax it to our Director of Bioinformatics at 301-838-0208. By signing below (or emailing from your own account), you agree to the terms above.

Name of the organism for which you want unpublished data (please type):

Name and job title:

Email address:

Physical mail address:

Upon receipt of this agreement, we will send you information on how to download the data from TIGR. This information will be sent to you using the email address you provide above.